

# Fennica linkitettynä datana tilannekatsaus

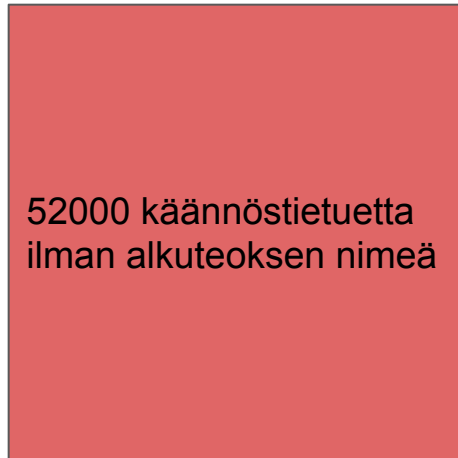
Osma Suominen

Tietomalliryhmän kokous 14.3.2017

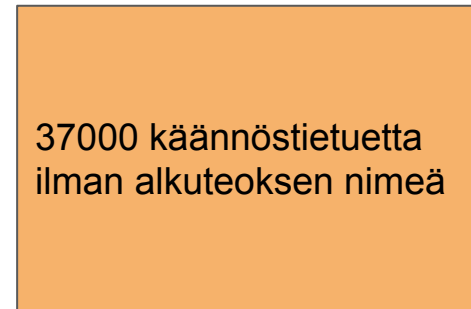
Tehtyä sitten viime näkemän

# Teosten eristäminen

Etsitään alkuteosten nimiä myös 130-kentästä ja 500-huomautuksista, jos 240 puuttuu.



**ennen**



**jälkeen**

# RDA tyypit

- RDA content type
- RDA carrier type
- RDA media type

esitetään nyt Schema.org datassa  
RDA Vocabularyn URI-tunnisteilla

# Organisaatiot ja kokoukset

- tekijänä (110/111/710/711)
- aiheena (610/611)
- julkaisijana (260b)

esitetään nyt Suomalaisten yhteisönimien  
URN-tunnisteella (jos löytyy)

# Henkilöiden synnyin- ja kuolinvuodet

Pekka Heikkisen näkemys: yli 25 vuotta kuolleilla ei tietosuojaaja

Poistetaan synnyin- ja kuolinvuositieto, jos

- henkilö on kuollut vuoden 1990 jälkeen, tai
- jos kuolinvuosi ei tiedossa, henkilö on syntynyt vuoden 1870 jälkeen

Tämän siivouksen jälkeen datassa ei pitäisi olla mitään arkaluontoista!

### # The original English language work

fennica:000215259work9 a schema:CreativeWork ;  
schema:about ysa:Y94527, ysa:Y96623, ysa:Y97136,  
ysa:Y97137, ysa:Y97575, ysa:Y99040,  
yso:p18360, yso:p19627, yso:p21034,  
yso:p2872, yso:p4403, yso:p9145 ;  
schema:author fennica:000215259person10 ;  
schema:inLanguage "en" ;  
schema:name "The illustrated A brief history of time" ;  
schema:workTranslation fennica:000215259 .

### # The Finnish translation (~expression in FRBR/RDA)

fennica:000215259 a schema:CreativeWork ;  
schema:about ysa:Y94527, ysa:Y96623, ysa:Y97136,  
ysa:Y97137, ysa:Y97575, ysa:Y99040,  
yso:p18360, yso:p19627, yso:p21034,  
yso:p2872, yso:p4403, yso:p9145 ;  
schema:author fennica:000215259person10 ;  
schema:contributor fennica:000215259person11 ;  
schema:inLanguage "fi" ;  
schema:name "Ajan lyhyt historia" ;  
schema:translationOfWork fennica:000215259work9 ;  
schema:workExample fennica:000215259instance26 ;  
[rdau:P60049](#) [rdacontent:1020](#) .

### # The manifestation (FRBR/RDA) / instance (BIBFRAME)

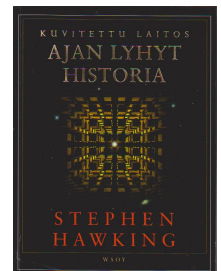
fennica:000215259instance26 a schema:Book, schema:CreativeWork ;  
schema:author fennica:000215259person10 ;  
schema:contributor fennica:000215259person11 ;  
schema:datePublished "2000" ;  
schema:description "Lisäpainokset: 4. p. 2002. - 5. p. 2005." ;  
schema:exampleOfWork fennica:000215259 ;  
schema:isbn "9510248215", "9789510248218" ;  
schema:name "Ajan lyhyt historia" ;  
schema:numberOfPages "248, 6 s. ." ;  
[rdau:P60048](#) [rdacarrier:1049](#) ;  
[rdau:P60050](#) [rdamedia:1007](#) ;  
[schema:publisher cn:32729A](#) . # WSOY

### # The original author

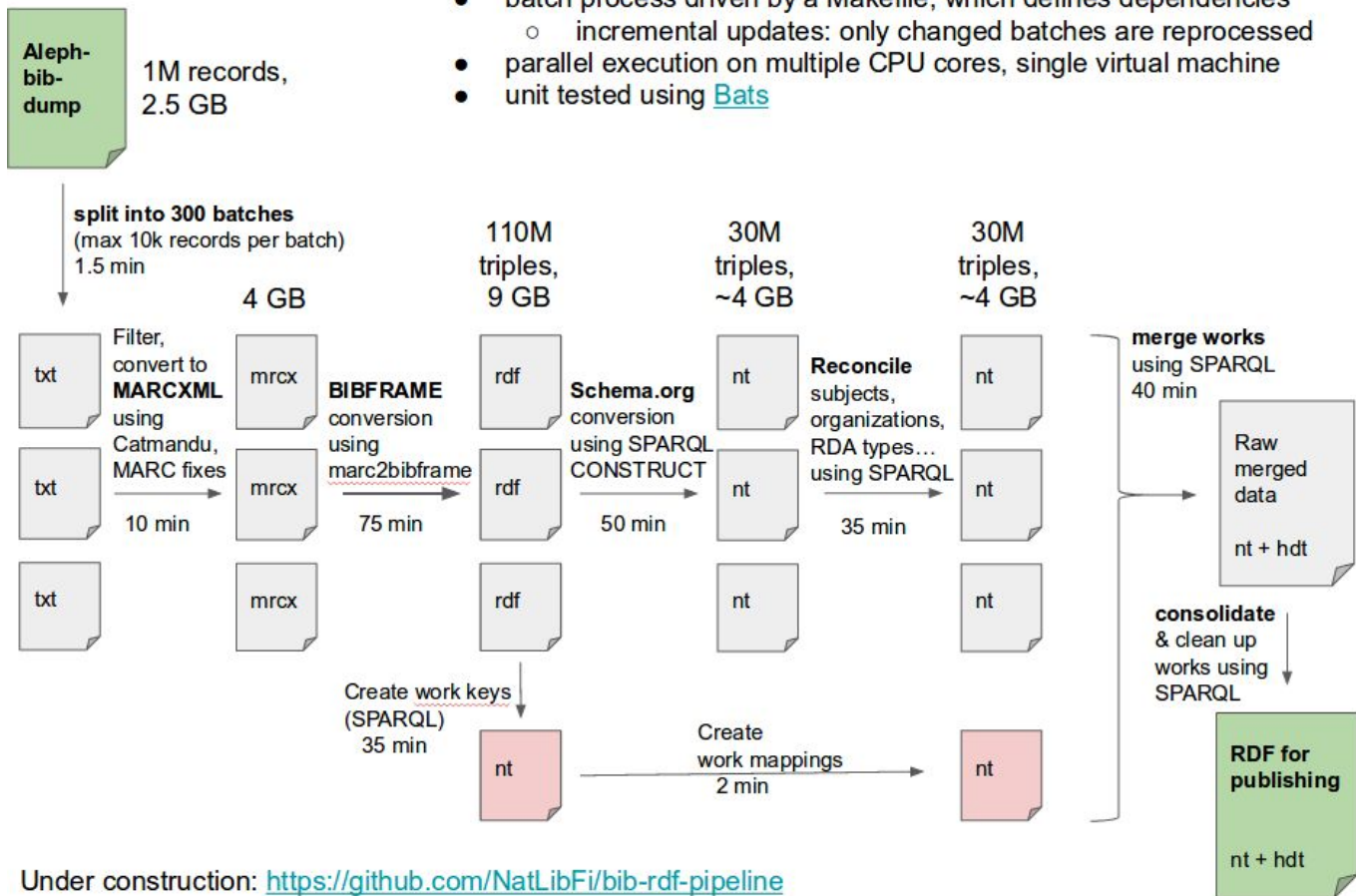
fennica:000215259person10 a schema:Person ;  
schema:name "Hawking, Stephen" .

### # The translator

fennica:000215259person11 a schema:Person ;  
schema:name "Varteva, Risto" .



# Fennica RDF conversion pipeline (draft)



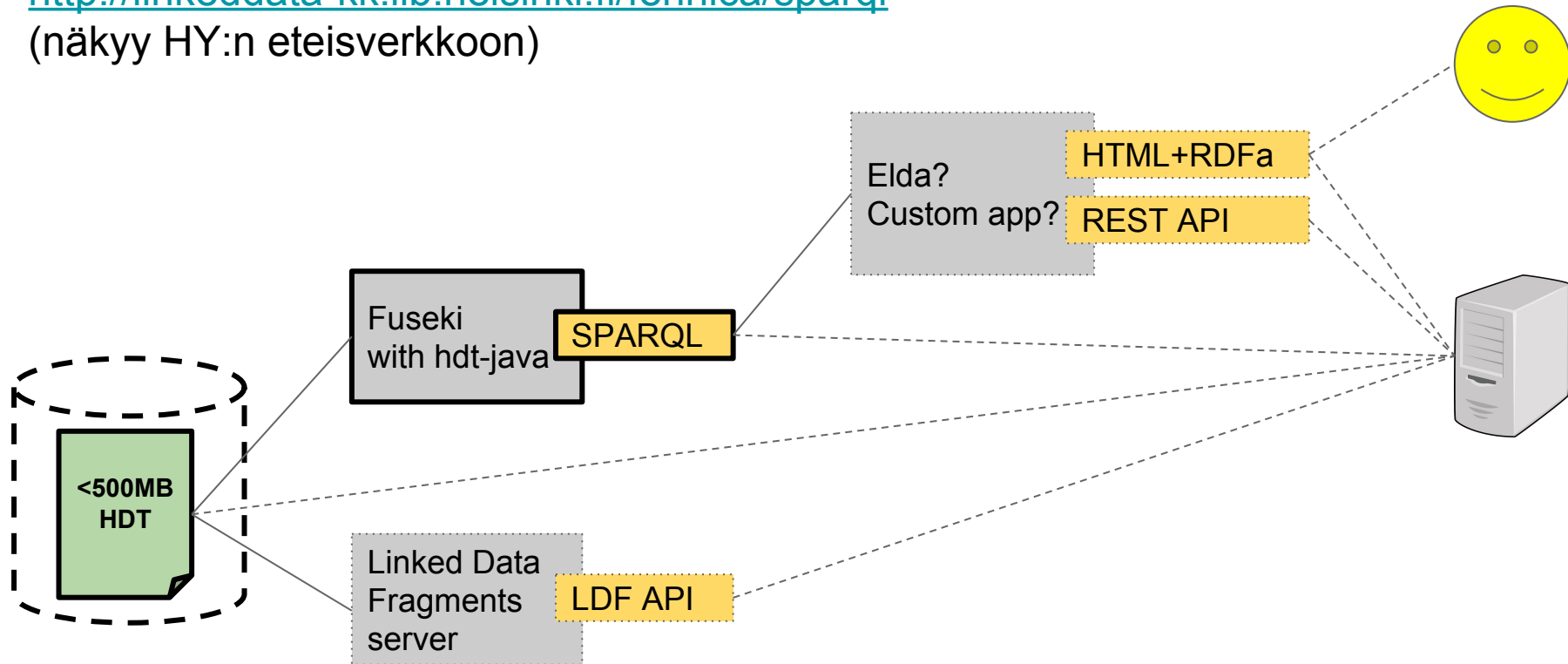
Under construction: <https://github.com/NatLibFi/bib-rdf-pipeline>



# SPARQL-rajapinta pystyssä!

<http://linkeddata-kk.lib.helsinki.fi/fennica/sparql>

(näkyä HY:n eteisverkkoon)



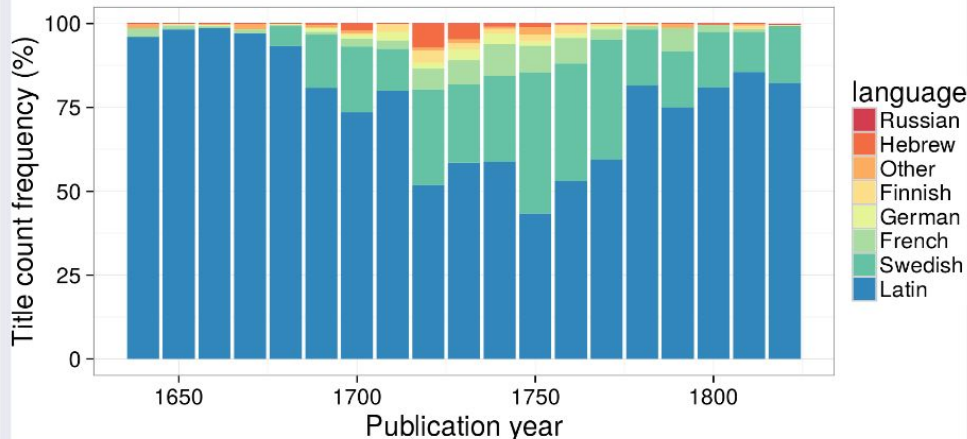
# Yhteistyö Mikko Tolosen DH-tutkimusryhmän kanssa

Annettu Tolosen ryhmälle

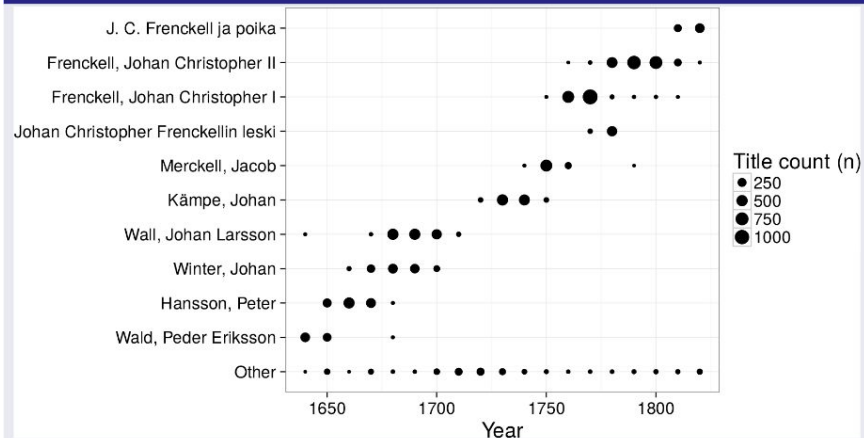
- MARC-tietueet (synt.vuodet siivottu)
- RDF-dumppi
- SPARQL-rajapinta

## Dissertation languages in Turku 1640-1828

### Dissertation languages (Fennica)



## Top publishers in Turku/Fennica

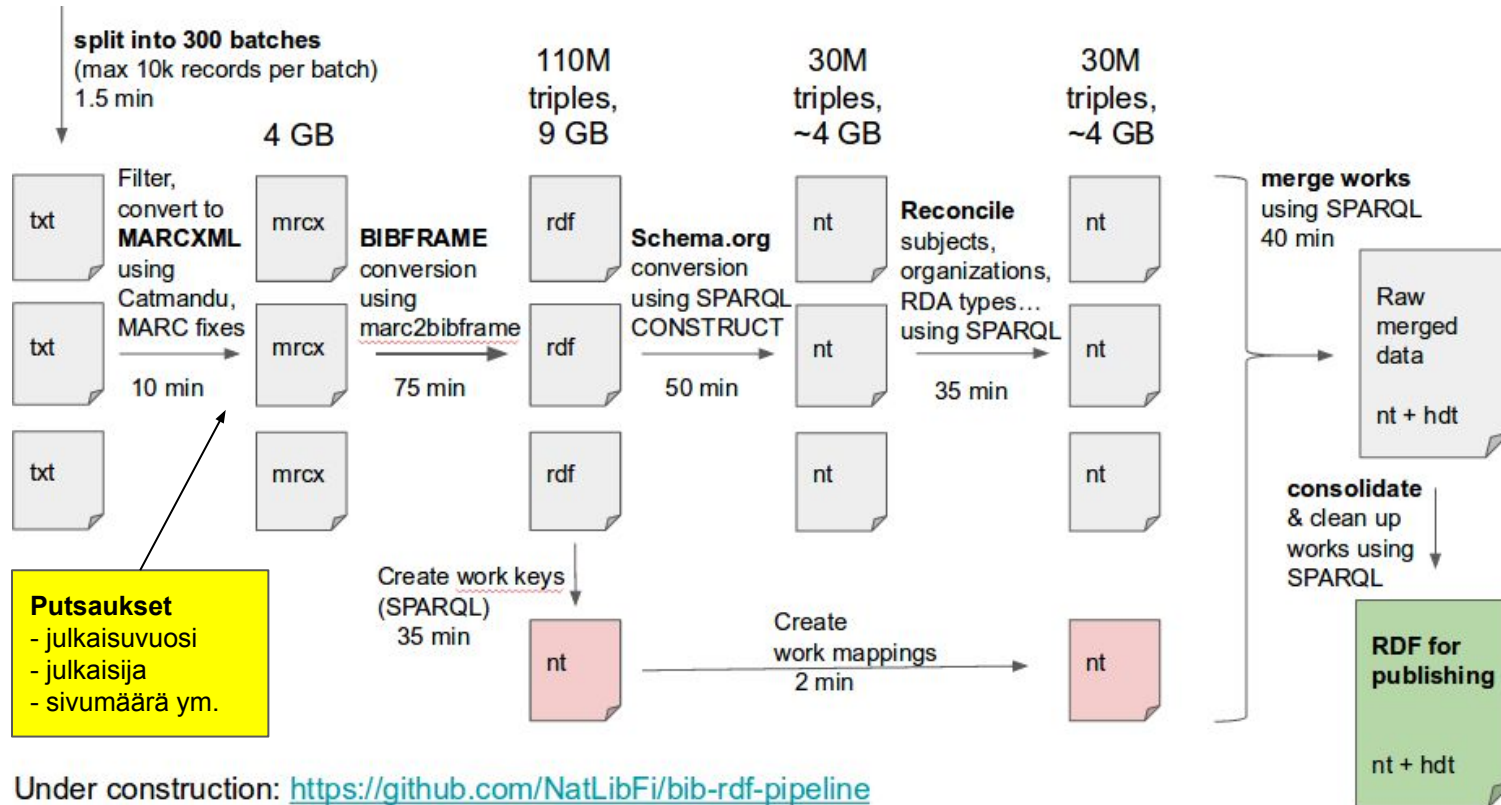


Lisää Tolosen ryhmän Fennica-analyyseja:

<https://github.com/rOpenGov/fennica/blob/master/inst/examples/20160715-Krakov-Fennica.pdf>

<http://www.ennenjanyt.net/2015/08/aatehistoria-ja-digitaalisten-aineistojen-mahdollisuudet/>

# Tolosen ryhmän putsaukset osaksi muunnosta



Under construction: <https://github.com/NatLibFi/bib-rdf-pipeline>

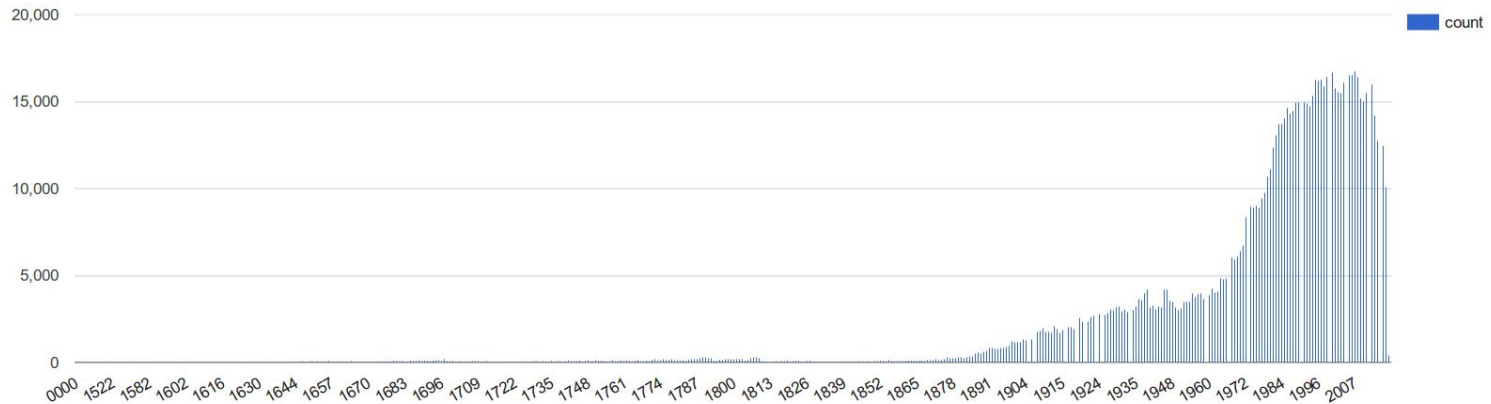
# Analyysi suoraan SPARQL-rajapinnan päällä

http://linkeddata-kk.lib.helsinki.fi/fennica/sparql

```
1 PREFIX schema: <http://schema.org/>
2 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
3 SELECT ?dateInt (COUNT(*) AS ?count)
4 WHERE {
5   ?inst schema:datePublished ?dateRaw .
6   BIND(REPLACE(?dateRaw, '\\[(.*)\\]', '$1') AS ?dateVal)
7   BIND(STRDT(?dateVal, xsd:nonNegativeInteger) as ?dateInt)
8   FILTER(isNumeric(?dateInt) && ?dateInt <= 2017)
9 }
10 GROUP BY ?dateInt
11 ORDER BY ?dateInt
12
```

Table Raw Response Pivot Table **Google Chart** Geo

Chart Config



# DCMI-webinaari 28.2.2017

## *From MARC silos to Linked Data silos? Data models for bibliographic Linked Data*

**Webinar Date:** Tuesday, 28 February 2017, 10:00am-11:15am EST (UTC 15:00 - World Clock: <http://bit.ly/suominen-2017>)

**Abstract:** Many libraries are experimenting with publishing their metadata as Linked Data to open up bibliographic silos, usually based on MARC records, to the Web. The libraries who have published Linked Data have all used different data models for structuring their bibliographic data. Some are using a FRBR-based model where Works, Expressions and Manifestations are represented separately. Others have chosen basic Dublin Core, dumbing down their data into a lowest common denominator format. The proliferation of data models limits the reusability of bibliographic data. In effect, libraries have moved from MARC silos to Linked Data silos of incompatible data models. There is currently no universal model for how to represent bibliographic metadata as Linked Data, even though many attempts for such a model have been made.

In this webinar, you'll see:

- a survey of published bibliographic Linked Data, the data models proposed for representing bibliographic data as RDF, and tools used for conversion from MARC records
- an analysis of different use cases for bibliographic Linked Data and how they affect the data model
- recommendations for choosing a data model

We also present efforts at the National Library of Finland to open up our bibliographic metadata, including the national bibliography Fennica, the national discography Viola and the article database Arto, as Linked Data while trying to learn from the examples of others. We are setting up a conversion process from MARC records to BIBFRAME and Schema.org compliant RDF, which we are going to publish as Linked Data using various technologies including a SPARQL endpoint, HDT compressed RDF dumps and a Linked Data Fragments API.

This webinar is an extended, in-depth version of the SWIB16 conference presentation "From MARC silos to Linked Data silos?"

**Minimum Participant Experience Level:** Basic familiarity of bibliographic metadata and Linked Data assumed

### **Presenter:**



**Osma Suominen** works as an information systems specialist at the National Library of Finland. His current activities are centered around the publishing of bibliographic data, including the Finnish national bibliography Fennica, as Linked Data. He is also one of the creators of the [Finto.fi](#) thesaurus and ontology service and is leading development of the Skosmos vocabulary browser used in Finto. Osma Suominen earned his doctoral degree at Aalto University while doing research on semantic portals and quality of controlled vocabularies within the FinnONTO series of projects. His past accomplishments include the Skosify vocabulary analysis and quality improvement tool, and [data.aalto.fi](#), the Linked Data service of Aalto University.

### **Youtube-tallenne:**

<https://www.youtube.com/watch?v=Ua9Wd90h97Q>

Seuraavia askeleita

# Uuden BIBFRAME-konversiotyökalun testaus

Library of Congress ja Index Data julkaisivat 13.3.2017 uuden MARC - BIBFRAME2 muunnostyökalun

<https://github.com/lcnetdev/marc2bibframe2>

“The Programs for converting the MARC data to BIBFRAME were written for the Library of Congress by Index Data. They are written in XSLT and are available for download on the Library of Congress Github site at <https://github.com/lcnetdev/marc2bibframe2> . The Library is currently working with these conversions in development of the Pilot project and we expect that they will be adjusted as work progresses.” -- Sally McCallum, BIBFRAME-lista 13.3.

Testattava miten tämä sopii nykyiseen muunnosputkeen vanhan marc2bibframe -konversiotyökalun tilalle.

# Henkilöauktoriteetin käyttö linkitettyssä datassa

Henkilöauktoriteettia ei tällä hetkellä oteta huomioon RDF-konversiossa.

Pitäisi...

- muuntaa henkilöauktoriteetti RDF:ksi ja julkaista se jossain (Finto?)
  - syntymä- ja kuolinvuodet siivottu edellä mainitulla logiikalla
  - muutenkin rajata julkaistavaa tietoa, jos on epäilyksiä - periaatteessa jo nimi ja tunniste riittää
- linkittää bibliografiset tiedot henkilöauktoriteettien tunnisteisiin RDF-datassa
  - odottaa sitä, että linkitys tehtäisiin Alephin/Melindan bib-tietueiden tasolla
- mielellään linkittää henkilöt ulkoisiin lähteisiin (VIAF, ISNI, Wikidata...)



# Datan käyttöliittymät

- Eldan (web-käyttöliittymä + REST-rajapinnat) pystytys SPARQL-rajapinnan varaan datan selailua varten
- Analyysiesimerkkien kehittäminen Tolosen ryhmän kanssa
  - SPARQL-kysely + visualisointi
  - toteutustekniikoina YASGUI ja SPARQL-kirjasto R-ohjelmointikielelle, jota tutkijat käyttävät

# Kansainvälisiä tapaamisia

1. LD4L Labs- ja LD4P-projektien yhteisötapaaminen 24-25.4.2017 Stanford
  - a. minut on kutsuttu mukaan, mahdollisuus pitää 5 min salamaesitys Fennica-konversiosta
2. “Days of Knowledge Organization”, 1.-2.6.2017 Oslo
  - a. minut on kutsuttu keynote-puhujaksi